

KOREAN PATENT ABSTRACTS

(11)Publication
number: **1020010057775 A**

(43)Date of publication of application:
05.07.2001

(21)Application number: **1019990061182** (71)Applicant: **KOREA ELECTRONICS &**

(22)Date of filing: **23.12.1999** (71)Applicant: **TELECOMMUNICATIONS
RESEARCH INSTITUTE**

(72)Inventor: **CHOI, SEUNG GWON**

(54) SYSTEM AND METHOD FOR AUTOMATICALLY PERFORMING TRANSLATION BASED ON SENTENCE FRAME

(57) Abstract:

PURPOSE: A system and method for automatically performing a translation based on a sentence frame is provided to naturally generate a translated sentence by removing a structural ambiguity on the basis of a sentence frame, which represents a frame of a sentence. CONSTITUTION: A source language morpheme analyzer(20) analyzes a morpheme of a source language sentence, and detects information on parts of speech of each word. A translation dictionary(21) stores a source vocabulary, a substitutive word of the vocabulary and quality information related to the word. A phrase unit syntax analyzer(22) represents an input sentence using the information detected from the source language morpheme analyzer(20). An original frame searcher(23) searches a constraint. An original frame database(24) stores a basic constraint of a relevant original frame. A sentence frame converter(25) converts the original into morpheme columns of an object language. A translation sentence frame database(26) stores constraint information of an original slot and designated information of a translation sentence slot. A slot translation database(27) stores constraint information of a part of speech of the original and designated information of a part of speech of the translation. An object language morpheme generator(28) analyzes the

morpheme of the object language.

(19) 대한민국특허청 (KR)
 (12) 공개특허공보 (A)

(51) . Int. Cl. 7
 G06F 17/28

(11) 공개번호 특2001-0057775
 (43) 공개일자 2001년07월05일

(21) 출원번호 10 - 1999 - 0061182
 (22) 출원일자 1999년12월23일

(71) 출원인 한국전자통신연구원
 오길록
 대전 유성구 가정동 161번지

(72) 발명자 심철민
 대전광역시서구만년동초원아파트110동602호
 여상화
 대전광역시유성구궁동395 - 3다솔아파트102동1105호
 정한민
 대전광역시유성구어은동1
 최승권
 대전광역시유성구어은동99한빛아파트113동1101호
 김영길
 대전광역시유성구어은동1
 서광준
 대전광역시유성구가정동161번지
 서영애
 대전광역시유성구송강동청솔아파트514동1109호
 김태완
 대전광역시유성구어은동한빛아파트105동605호
 박상규
 대전광역시유성구송강동한마을아파트105동1305호
 박세영
 대전광역시유성구어은동한빛아파트133동1004호

(74) 대리인 특허법인 신성 박혜천
 특허법인 신성 원석희
 특허법인 신성 최종식
 특허법인 신성 박정후
 특허법인 신성 정지원

심사청구 : 있음

(54) 문장골격을 기반으로 한 자동번역 시스템 및 그 방법

요약

본 발명은 문장골격을 기반으로 한 자동번역 시스템 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체에 관한 것으로, 문장의 골격을 표현하고 있는 문틀에 기반하여, 구조적인 모호성을 해소한 자연스러운 번역문장을 생성할 수 있는 문장골격을 기반으로 한 자동번역 시스템 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공하기 위하여, 원시언어 형태소 분석기 및 구 단위 구문분석기로부터 구 단위 구분분석 결과를 수신하여 원문틀 데이터베이스에 저장되어 있는 제약조건을 매칭시키는 제1 단계; 상기 제약조건이 매칭된 구문에 슬롯 단위의 어순변환 및 슬롯 내부의 품사단위의 어순변환을 수행하는 제2 단계; 및 상기 슬롯 단위 및 품사단위로 어순 변환된 원시문장을 정규화된 자질값을 갖는 목적언어 형태소로 구성하는 제3 단계를 포함하며, 자동번역 시스템 등에 이용됨.

대표도

도 2

색인어

자동번역, 문장골격, 슬롯단위, 제약조건, 가중치, 자질값

명세서

도면의 간단한 설명

도 1 은 종래의 자동번역 시스템의 일실시에 구성도.

도 2 는 본 발명에 따른 자동번역 시스템의 일실시에 구성도.

도 3 은 본 발명에 따른 원문틀 탐색기 및 구문 변환기의 일실시에 구성도.

도 4 는 본 발명에 따른 문장골격을 기반으로 한 자동번역 방법에 대한 일실시에 처리 흐름도.

* 도면의 주요 부분에 대한 부호의 설명

20 : 원시언어 형태소 분석기 21 : 번역사전

22 : 구 단위 구분분석기 23 : 원문틀 탐색기

24 : 원문틀 데이터베이스 25 : 문틀 변환기

26 : 대역문틀 데이터베이스 27 : 슬롯대역 데이터베이스

28 : 목적언어 형태소 생성기

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 자동번역 시스템 및 그 방법에 관한 것으로, 특히 입력된 원시언어 문장으로부터 문장의 골격이라 할 수 있는 문틀(sentence frame)을 파악하고, 이를 바탕을 대역문의 문틀로 변환한 후, 대역어를 생성해 낼 수 있는 문장골격을 기반으로 한 자동번역 시스템 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체에 관한 것이다.

도 1에 도시된 바와 같은 종래의 자동번역 시스템은, 원시언어 형태소 분석기(10)에서 변역사전(14)을 이용하여 어휘의 품사를 파악한 후, 구문분석기(11)가 구문분석 규칙(15)을 이용하여 문장의 통사적인 수식구조를 파악한다.

입력문장의 구조분석이 완료되면, 문장의 통사구조를 트리구조로 표현하여 목적언어의 구조로 변환해 주는 변환기(12)로 전달한다. 상기 변환기(12)는 변환규칙(16) 및 변역사전(14)을 참조하여 이용하여 원시언어 트리를 목적언어 트리로 변환하고, 목적언어 형태소 생성기(13)는 목적언어로 변환된 트리를 탐색하여, 해당 노드의 형태소를 생성하며, 최종적인 번역문장을 생성한다. 여기서, 문장의 통사구조를 파악하는 구문분석은, 품사단위의 어휘정보와 상기 원시언어 형태소 분석기(10)에서 제공하는 자질정보등의 국부적인 정보만을 고려하여 해당 단어들을 구 단위로 결합하고자 하므로, 여러 가지 모호성(ambiguity)이 발생할 수 있다.

일반적으로, 상기한 바와 같은 구조적인 모호성은 구문분석 단계에서 적용되는 규칙의 선호도 값들을 이용하여 전체 구문트리의 신뢰도를 측정하여 해소하나, 영어의 경우 20 단어를 넘어가는 문장의 경우, 구문분석시 발생하는 모호성이 증가함으로써, 구문분석의 정확도가 현저하게 떨어진다.

상기 구문분석의 모호성을 해소하기 위해 단순히 품사단위의 결합규칙만으로 구조를 분석하는 것이 아니라, 어휘와 품사의 혼합으로써 구문분석 규칙을 기술하는 방법이 있다.

그러나, 상기한 바와 같이 어휘와 품사의 혼합으로써 구문분석 규칙을 서술하는 방법은, 구문분석시 임의의 어휘에 대해 품사규칙 뿐만 아니라, 어휘규칙까지 모두 적용해야 하므로, 구문분석의 속도가 느리고, 메모리를 많이 필요로 하며, 국부적인 정보만을 이용하므로 모호성의 문제를 근본적으로 해결할 수 없는 문제가 있다.

상기 구문분석의 모호성을 해소하기 위한 다른 방법으로, 대용량의 번역 예문으로부터 단편적인 번역 예제를 추출하고, 각 예제마다 해당하는 대역구조를 명시한 후, 실제 번역시에는 입력 문장에 대해 부분적인 예제들을 탐색하여 해당하는 대역구조를 변환하는 방법이 있다.

그러나, 상기한 바와 같은 구문분석의 모호성을 해소하기 위한 방법은, 어휘단위의 대용량의 예제를 전체로 함으로써, 일정한 수준의 번역 품질을 내기 위해 구축해야 하는 번역지식의 양이 방대하며, 예제의 구축에 일정한 기준이 없어, 번역지식의 튜닝이 어렵고, 예제 역시 부분적인 번역정보만을 표현하고 있으므로, 모호성의 문제를 해결하지 못하는 문제가 있다.

결론적으로, 종래의 자동번역 시스템에서의 구문변환 방법은, 국부적인 정보만을 사용함으로써, 구조적인 모호성을 해소할 수 없는 문제가 있다.

발명이 이루고자 하는 기술적 과제

따라서, 본 발명은 상기와 같은 문제점을 해결하기 위해 안출된 것으로, 문장의 골격을 표현하고 있는 문틀에 기반하여, 구조적인 모호성을 해소한 자연스러운 번역문장을 생성할 수 있는 문장골격을 기반으로 한 자동번역 시스템 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공하는데 그 목적이 있다.

발명의 구성 및 작용

상기 목적을 달성하기 위한 본 발명의 시스템은, 자동번역 시스템에 있어서, 원시언어 문장의 형태소 분석을 수행하여

각 단어의 품사정보를 추출하는 원시언어 형태소 분석수단; 원시 어휘와 그에 대한 대역어 및 대역어와 관련된 자질정보를 저장하는 자질정보 저장수단; 상기 원시언어 형태소 분석수단에서 추출된 품사정보를 이용하여 입력문장을 슬롯의 열로 표현하는 구 단위 구문분석수단; 상기 원시언어 문장의 슬롯 열을 바탕으로 제약조건을 탐색하는 원문틀 탐색수단; 해당하는 원문틀의 기본적인 제약조건을 저장하는 원문틀 정보 저장수단; 상기 원문틀 탐색수단으로부터 제약조건이 매칭된 원문을 수신하여 정규화된 자질값을 갖는 목적언어의 형태소열로 변환하는 문틀 변환수단; 원문 슬롯의 제약정보 및 대역문 슬롯의 지정정보를 저장하는 대역문틀 정보 저장수단; 원문 품사의 제약정보 및 대역문 품사의 지정정보를 저장하는 슬롯대역 정보 저장수단; 및 상기 문틀 변환수단으로부터 목적언어의 형태소 열을 수신하여 목적언어의 형태소를 분석하는 목적언어 형태소 생성수단을 포함한다.

한편, 본 발명의 방법은, 자동번역 시스템에 적용되는 문장골격을 기반으로한 자동번역 방법에 있어서, 원시언어 형태소 분석기 및 구 단위 구문분석기로부터 구 단위 구문분석 결과를 수신하여 원문틀 데이터베이스에 저장되어 있는 제약조건을 매칭시키는 제 1 단계; 상기 제약조건이 매칭된 구문에 슬롯 단위의 어순변환 및 슬롯 내부의 품사단위의 어순변환을 수행하는 제 2 단계; 및 상기 슬롯 단위 및 품사단위로 어순 변환된 원시문장을 정규화된 자질값을 갖는 목적언어 형태소로 구성하는 제 3 단계를 포함한다.

또한, 본 발명은, 마이크로 프로세서를 구비한 자동번역 시스템에, 원시언어 형태소 분석기 및 구 단위 구문분석기로부터 구 단위 구문분석 결과를 수신하여 원문틀 데이터베이스에 저장되어 있는 제약조건을 매칭시키는 제 1 기능; 상기 제약조건이 매칭된 구문에 슬롯 단위의 어순변환 및 슬롯 내부의 품사단위의 어순변환을 수행하는 제 2 기능; 및 상기 슬롯 단위 및 품사단위로 어순 변환된 원시문장을 정규화된 자질값을 갖는 목적언어 형태소로 구성하는 제 3 기능을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공한다.

상술한 목적, 특징들 및 장점은 첨부된 도면과 관련한 다음의 상세한 설명을 통하여 보다 분명해 질 것이다. 이하, 첨부된 도면을 참조하여 본 발명에 따른 바람직한 일실시예를 상세히 설명한다.

도 2 는 본 발명에 따른 자동번역 시스템의 일실시예 구성도이다.

도면에 도시된 바와 같이, 본 발명에 따른 자동번역 시스템은, 원시언어 문장의 형태소 분석을 수행하여 각 단어의 품사정보를 추출하는 원시언어 형태소 분석기(20)와, 원시 어휘와 그에 대한 대역어 및 대역어와 관련된 자질정보를 저장하는 번역사전(21)과, 상기 원시언어 형태소 분석기(20)에서 추출된 품사정보를 이용하여 입력문장을 슬롯의 열로서 표현하는 구 단위 구문분석기(22)와, 원시언어 문장의 슬롯 열을 바탕으로 제약조건을 탐색하는 원문틀 탐색기(23)와, 해당하는 원문틀의 기본적인 제약조건을 저장하는 원문틀 데이터베이스(24)와, 상기 원문틀 탐색기(23)로부터 제약조건이 매칭된 원문을 정규화된 자질값을 갖는 목적언어의 형태소열로 변환하는 문틀 변환기(25)와, 원문 슬롯의 제약정보 및 대역문 슬롯의 지정정보를 저장하는 대역문틀 데이터베이스(26)와, 원문 품사의 제약정보 및 대역문 품사의 지정정보를 저장하는 슬롯대역 데이터베이스(27)와, 상기 문틀 변환기(25)로부터 정규화된 자질값을 수신하여 목적언어의 형태소를 분석하는 목적언어 형태소 생성기(28)를 포함한다.

그리고, 상기 구 단위 구문분석기(22)가 표현하는 슬롯은, 문장의 골격을 좌우하는 특수한 품사(예를 들면, 동사, 기호, 접속사 및 관계사)는 슬롯으로 연결하지 않고 별도의 노드로 구성한다. 상기 품사들은 하나의 형태소가 하나의 슬롯을 구성하며, 이외의 형용사, 부사, 관사, 명사 및 전치사 등의 품사들은 서로 연결되어 슬롯을 구성할 수 있다.

상기 구 단위 구문분석기(22)가 원시문장을 슬롯으로 표현하는 예를 설명하기로 한다.

예를 들어 "The government unveils financial market stabilization measures which include the injection of public funds into ailing financial institutions." 와 같은 문장의 경우, 형태소 분석결과는 다음과 같다.

(어휘/품사/기본형/세부품사)

the/DET/the/DT

government/NOUN/government/NN

unveils/VERB/unveil/VBZ

financial_market/NOUN/financial_market/JJ

stabilization_measures/NOUN/stabilization_measure/NN

which/CONJ/which/CONJ

include/VERB/include/VB

the/DET/the/DT

injection/NOUN/injection/NN

of/PREP/of/IN

public_funds/NOUN/public_fund/JJ

into/PREP/into/IN

ailing_financial_institutions/NOUN/ailing_financial_institution/JJ

PERIOD/PUNCT/PERIOD

상기한 바와 같은 결과를 상기 구 단위 구문분석기(22)를 이용하여 슬롯 열로 연결한 결과는 다음과 같다.

(슬롯명/어휘)

NP1/the_government

VERB1/unveils

NP2/financial_market_stabilization_measures

CONJ1/which

VERB2/include

NP3/the_injection_of_public_funds_into_ailing_financial_institutions

상기한 바와 같은 슬롯 열을 원문틀로써 표현한 예는 다음과 같다. 원문들은 원시문장의 구 단위 구문분석 결과인 슬롯 열을 기호화한 것이다.

(원문틀 탐색기)

NP VERB NP CONJ VERB NP

상기한 바와 같은 원문틀 탐색키를 이용하여 상기 원문틀 탐색기(23)는 상기 원문틀 데이터베이스(24)를 탐색한다. 원문틀 접근과정은 상기 원문틀 데이터베이스(24)에 저장된 대역문틀의 기본 제약조건을 만족하는 경우, 해당 대역문틀의 실제 내용을 접근하는 단계로 구성된다. 이와 같이, 원문틀과 대역문틀 사이에 제약조건을 설정한 이유는 첫째, 문장의 전체적인 골격을 나타내는 원문틀에서 직접 대역문틀로 접근할 경우, 동일한 원문틀에 대해 다양한 대역문틀이 가능하므로 처리흐름상 하나의 키에 저장된 데이터베이스 내용의 크기가 커지고, 처리시간이 증가하며, 둘째, 대용량의 번역지식 구축시, 원문틀과 대역문틀을 직접 연결할 경우, 동일한 원문틀에 대해 대역문틀을 구별하여 구축하기 위하여 고려해야 할 후보지식의 수가 많고, 그들간의 차이점을 쉽게 파악할 수 없기 때문이다.

첨부된 도 3은 본 발명에 따른 원문틀 탐색기 및 문틀 변환기의 일실시에 상세구성도이다.

상기 원문틀 탐색기(23)는, 상기 구 단위 구문분석기(22)로부터 원시문장의 슬롯 열을 입력받아 상기 원문틀 데이터베이스(24)에서 해당하는 원문틀의 제약조건을 탐색하여 수신된 원시문장 슬롯열의 통사적인 특징을 파악하는 원문틀 제약 조건 검사기(30)를 포함한다.

상기 원문틀 탐색기(23)에서 원문틀 탐색에 성공하면, 해당 원문틀의 통사적인 특징을 파악하기 위하여 상기 원문틀 제약조건 검사기(30)가 상기 원문틀 데이터베이스(24)의 제약조건 부분을 탐색하여, 제약조건을 매칭한다. 그리고, 각 제약조건마다 부가되어 있는 상기 대역문틀 데이터베이스(26)의 탐색키를 이용하여 해당 대역문틀에 접근한다.

다음의 상기 원문틀 데이터베이스(24)에 저장된 제약조건의 예이다.

{NP VERB:[vb,11] NP CONJ:[rel] VERB:[vb,t1] NP} 대역문틀 내용으로의 포인터1, 대역문틀 내용으로의 포인터2, ...

상기 제약조건의 예에서, VERB:[vb,t1]의 경우, 슬롯의 명칭은 VERB이며, 제약조건으로써 동사의 활용형태는 현재형(vb)이고, 동사의 형식은 타동사(t1)라는 의미이다. 상기와 같은 제약조건은, 명확한 구축지침을 제시함으로써, 대용량의 예문을 일관성 있게 구축할 수 있다.

그리고, 상기 문틀 변환기(25)는, 입력된 대역문에 대해 원문의 각 슬롯별로 해당하는 자질 제약조건 스크립트를 실행하여 만족하는가를 검사하고 슬롯순서를 정렬하며 변환된 각각의 대역문 슬롯에 대해 자질값 지정 스크립트를 수행하여 자질값을 부여한 후 목적언어 형태로 구조변환을 수행하는 대역문틀 분석기(31)와, 상기 대역문틀 분석기(31)로부터 구조 변환된 결과를 수신하고 상기 슬롯대역 데이터베이스(27)를 참조하여 각 슬롯별로 내부의 형태소들에 대한 어순변환 및 해당하는 자질값을 지정하는 2단계 구조변환을 수행하는 슬롯대역 구조분석기(32)와, 상기 번역사전(21)을 참조하여 각 어휘별로 대역어를 추출하는 어휘 변환기(33)와, 상기 대역문틀 분석기(31)에서 계산된 가중치 값을 이용하여 가장 적절한 대역문들을 선택하는 문틀 선택기(34)와, 구조변환 과정에서 지정된 다수의 자질값들을 정규화하여 상기 목적언어 형태소 생성기(28)에 전송하는 자질 정련기(35)를 포함한다.

상기 원문틀 제약조건 검사기(30)에서 원문틀 제약조건 매칭에 성공한 후, 해당하는 제약조건에 부가된 상기 대역문틀 데이터베이스(26)로의 포인터 정보를 이용하여, 상기 대역문틀 분석기(31)에서 상기 대역문틀 데이터베이스(26)에 접근한다. 상기 대역문틀 데이터베이스(26)에는 다음과 같이 원문 제약정보 및 대역문 지정정보가 저장되어 있다.

{원문슬롯1:[자질 제약조건 스크립트] 원문슬롯2:[자질 제약조건 스크립트]...} ->

{대역문슬롯1:[자질값 지정 스크립트] 대역문슬롯1:[자질값 지정 스크립트]...}

상기 원문 제약정보는, 원문틀에 기술된 각 원문 슬롯별로 특별한 제약조건들을 스크립트 언어(script language) 형태로 기술할 수 있다. 예를 들어, VERB1:[errot=="[have]_AND" etype=="[t1]]과" 같은 경우는, 원문 슬롯명이 VERB이고, 어휘의 기본형이 "have"이고, 동사의 형식이 타동사(t1)인 경우를 의미한다.

상기 대역문틀 분석기(31)는 먼저 하나의 대역문틀에 대해 원문의 각 슬롯별로 해당하는 자질 제약조건 스크립트를 실행하여 만족하는지를 검사한 후, 대역문의 각 슬롯별로 슬롯 순서를 정렬한다. 상기 예에서 원문 슬롯 순서와 대역문의 슬롯 순서를 다르게 배열한다. 예를 들어, "I love you"라는 입력문자의 경우, 원문은 "NP1 VERB1 NP2"와 같고, 이에 대응하는 대역문은 "NP1 NP2 VERB1"와 같이 표현되어 어순변환을 수행한다.

이와 같이, 어순변환을 수행한 후, 변환된 대역문 슬롯들 각각에 대해 자질값 지정 스크립트를 실행하여 특정 자질값을 지정한다. 예를 들어, "I love you"의 경우, "NP1 NP2 VERB1"과 같이 대역 슬롯의 어순을 변환한 후, NP1의 격 정보로는 "주격"을, NP2의 격 정보로는 "목적격"을 지정한다.

상기 입력예문 "The government unveils financial market stabilization measures which include the injection of public funds into ailing financial institutions."에 대한 대역문틀은 다음과 같다.

(슬롯명/어휘/자질정보)

NP1/the_government/주제격

NP3/the_injection_of_public_funds_into_ailing_financial_institutions/목적격

VERB2/include/평서형

CONJ1/which/연결형

NP2/financial_market_stabilization_measures/주격

VERB1/unveils/평서형

입력문장의 형태소 분석에서 발생하는 모호성과 구 단위 구문분석에서 발생하는 모호성 및 상기 번역사전(21)에 다수의 정보가 기술되어 있음으로 발생하는 모호성으로 인해 하나의 입력문장에 대해 다수의 원문틀이 매칭될 수 있으므로, 다수의 대역문틀이 후보로서 선택될 수 있다. 상기 대역문틀 후보 중에서 원문의 자질 제약조건 스크립트를 만족하는 정도에 따라 가중치를 부여한다.

상기 슬롯대역 데이터베이스(27)의 내용구성은 다음과 같다.

{원문품사1:[자질 제약조건 스크립트] 원문품사2:[자질 제약조건 스크립트]...} ->

{대역문품사1:[자질값 지정 스크립트] 대역문품사2:[자질값 지정 스크립트]...}

상기한 바와 같이, 상기 슬롯대역 데이터베이스(27)의 내용구성은 상기 대역문틀 데이터베이스(26)의 내용구성과 유사하나, 상기 대역문틀 데이터베이스(26)의 적용대상은 슬롯단위이고, 상기 슬롯대역 데이터베이스(27)의 적용대상은 슬롯 내부의 품사단위라는 점에 차이가 있다.

입력 예문 "The government unveils financial market stabilization measures which include the injection of public funds into ailing financial institutions."에 대한 대역문틀 중 두 번째 슬롯인 NP3(the_injection_of_public_funds_into_ailing_financial_institutions)에 대한 상기 슬롯대역 데이터베이스(27)의 예는 다음과 같다.

(NP3에 해당하는 슬롯대역의 예)

{DET1 NOUN1 PREP1 NOUN2 PREP2 NOUN3} -> {NOUN3 PREP1 NOUN2 PREP2 NOUN1}

이제 첨부된 도 4 를 참조하여 본 발명에 따른 문장골격을 기반으로한 자동번역 방법에 대해 상세히 설명하면 다음과 같다.

도 4 는 본 발명에 따른 문장골격을 기반으로한 자동번역 방법에 대한 일실시에 처리 흐름도이다.

먼저, 입력문장에 대해 원시언어 형태소 분석기(20)와 구 단위 구문분석기(22)를 거쳐 생성된 구 단위 구문분석 결과를 메모리 장치로부터 적재하고(400), 상기 구 단위 구문분석 결과가 올바른 결과인가를 판단한다(402).

상기 판단결과(402), 올바른 결과가 아니면, 자동번역 작업을 종료하고, 올바른 결과이면, 상기 적재된 구 단위 구문분석 결과를 이용하여 원문틀 탐색기를 생성하고(404), 상기 원문틀 탐색기를 이용하여 원문틀 데이터베이스(24)를 탐색한 후(406), 상기 원문틀 데이터베이스(24) 탐색결과가 올바른가를 판단한다(408).

상기 판단결과(408), 올바른 탐색결과가 아니면, 자동번역 작업을 종료하고, 올바른 탐색결과이면, 상기 원문틀 데이터베이스(24)에서 탐색한 제약조건을 입력문장에 매칭하고(410), 제약조건 매칭에 성공한 후보가 존재하는지를 판단한다(412).

상기 판단결과(412), 제약조건 매칭에 성공한 후보가 존재하지 않으면, 자동번역 작업을 종료하고, 제약조건 매칭에 성공한 후보가 존재하면, 상기 제약조건에 부가되어 있는 대역문틀 접근 포인터를 이용하여 대역문틀 데이터베이스(26)를 탐색하고(414), 상기 대역문틀 데이터베이스(26)에서 탐색된 대역문틀 정보를 적용한다(416). 여기서, 상기 대역문틀 정보 적용과정은, 원문 슬롯의 제약조건 스크립트를 실행하여 가중치를 계산하고, 상기 원문 슬롯과 대역문 슬롯의 위치를 고려하여 문장을 구성하는 슬롯단위의 어순변환을 수행한 후, 각각의 대역문 슬롯에 대해 자질 지정 스크립트를 수행하여 해당하는 자질값을 지정한다.

그리고, 상기 과정을 통해 어순이 변환된 각 슬롯들에 대해 슬롯대역 데이터베이스(27)를 탐색하여(418), 슬롯대역 정보를 적재하고, 상기 대역문틀 적용에 성공한 모든 후보 대역문틀에 대해 각각의 문틀을 구성하는 슬롯별로 상기 슬롯대역 데이터베이스(27)의 슬롯대역 정보를 적용한다(420). 여기서, 상기 슬롯대역 정보 적용과정은, 원문 품사에 대한 제약조건 스크립트를 수행하고, 상기 제약조건 스크립트를 만족하는 원문 품사 및 대역문 품사열의 위치를 고려하여 슬롯 내부의 품사단위의 어순변환을 수행하고, 대역문부의 해당 품사에 기술된 자질 지정 스크립트를 실행하여 해당하는 자질을 지정한다.

상기 슬롯대역 정보의 적용이 완료되면, 번역사전(21)을 이용하여 원시언어 단어에 해당하는 목적언어 단어로 어휘변환을 수행하고(422), 상기 가중치 값을 이용하여 최종번역 결과로서 하나의 대역문틀을 선택한 후(424), 정규화된 자질값을 갖는 목적언어 형태소로 구성한다(426).

이상에서 설명한 본 발명은, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 있어 본 발명의 기술적 사상을 벗어나지 않는 범위 내에서 여러 가지 치환, 변형 및 변경이 가능하므로 전술한 실시예 및 첨부된 도면에 한정되는 것이 아니다.

발명의 효과

상기와 같은 본 발명은, 문장단위의 구문분석을 수행하지 않고, 입력문장을 구 단위 슬롯으로 표현하고, 이에 대한 구조변환 규칙을 데이터베이스화하여 적용함으로써, 번역을 위한 번역지식 구축의 일관성을 유지할 수 있고, 대용량의 번역지식을 구축할 수 있으며, 구문분석 단계의 모호성으로 인한 번역품질의 저하를 극복할 수 있는 효과가 있다.

(57) 청구의 범위

청구항 1.

자동번역 시스템에 있어서,

원시 어휘와 그에 대한 대역어 및 대역어와 관련된 자질정보를 저장하는 자질정보 저장수단;

원시언어 문장의 형태소 분석을 수행하여 각 단어의 품사정보를 추출하는 원시언어 형태소 분석수단;

상기 원시언어 형태소 분석수단에서 추출된 품사정보를 이용하여 입력문장을 슬롯의 열로 표현하는 구 단위 구문분석 수단;

해당하는 원문틀의 기본적인 제약조건을 저장하는 원문틀 정보 저장수단;

상기 구 단위 구문분석수단으로부터 수신한 원시언어 문장의 슬롯 열을 바탕으로 상기 원문틀 정보 저장수단에서 제약 조건을 탐색하는 원문틀 탐색수단;

원문 슬롯의 제약정보 및 대역문 슬롯의 지정정보를 저장하는 대역문틀 정보 저장수단;

원문 품사의 제약정보 및 대역문 품사의 지정정보를 저장하는 슬롯대역 정보 저장수단;

상기 원문틀 탐색수단으로부터 수신한 제약조건이 매칭된 원문을 상기 대역문틀 정보 저장수단 및 슬롯대역 정보 저장 수단을 참고하여 정규화된 자질값을 갖는 목적언어의 형태소열로 변환하는 문틀 변환수단; 및

상기 문틀 변환수단으로부터 목적언어의 형태소 열을 수신하여 목적언어의 형태소를 분석하는 목적언어 형태소 생성수 단

을 포함하는 문장골격을 기반으로 한 자동번역 시스템.

청구항 2.

제 1 항에 있어서,

상기 원문틀 탐색수단은,

상기 구 단위 구문분석수단으로부터 원시문장의 슬롯 열을 입력받아 상기 원문틀 정보 저장수단에서 해당하는 원문틀의 제약조건을 탐색하여 상기 원시문장 슬롯열의 통사적인 특징을 파악하는 원문틀 제약조건 검사수단을 포함하는 문장골격을 기반으로 한 자동번역 시스템.

청구항 3.

제 1 항 또는 제 2 항에 있어서,

상기 문틀 변환수단은,

상기 원문틀 제약조건 검사수단으로부터 입력된 대역문에 대해 자질 제약조건 스크립트, 슬롯순서를 정렬, 자질값 부여 및 목적언어 형태로 구조변환을 수행하는 대역문틀 분석수단;

상기 대역문틀 분석수단으로부터 구조 변환된 결과를 수신하고, 상기 슬롯대역 정보 저장수단을 참조하여 각 슬롯별로 내부의 형태소들에 대한 어순변환 및 해당하는 자질값을 지정하는 슬롯대역 구조분석수단;

상기 자질정보 저장수단을 참조하여 각 어휘별로 대역어를 추출하는 어휘 변환수단;

상기 대역문틀 분석수단에서 계산된 가중치 값을 이용하여 가장 적절한 대역문틀을 선택하는 문틀 선택수단; 및

상기 대역문틀 분석수단에서 구조변환 결과를 수신하고, 지정된 다수의 자질값들을 정규화하여 상기 목적언어 형태소 생성수단으로 전송하는 자질 정련수단을 포함하는 문장골격을 기반으로 한 자동번역 시스템.

청구항 4.

자동번역 시스템에 적용되는 문장골격을 기반으로 한 자동번역 방법에 있어서,

원시언어 형태소 분석기 및 구 단위 구문분석기로부터 구 단위 구분분석 결과를 수신하여 원문틀 데이터베이스에 저장되어 있는 제약조건을 매칭시키는 제 1 단계;

상기 제약조건이 매칭된 구문에 슬롯 단위의 어순변환 및 슬롯 내부의 품사단위의 어순변환을 수행하는 제 2 단계; 및

상기 슬롯 단위 및 품사단위로 어순 변환된 원시문장을 정규화된 자질값을 갖는 목적언어 형태소로 구성하는 제 3 단계를 포함하는 문장골격을 기반으로 한 자동번역 방법.

청구항 5.

제 4 항에 있어서,

상기 제 1 단계는,

입력문장에 대해 상기 원시언어 형태소 분석기 및 상기 구 단위 구문분석기를 거쳐 생성된 구 단위 구분분석 결과를 적재하는 제 4 단계;

상기 구 단위 구분분석 결과가 올바른 결과인가를 확인하여, 상기 적재된 구 단위 구분분석 결과를 이용하여 원문틀 탐색키를 생성하는 제 5 단계;

상기 원문틀 탐색키를 이용하여 상기 원문틀 데이터베이스를 탐색하는 제 6 단계; 및

상기 원문틀 데이터베이스의 탐색결과가 올바른가를 확인하여, 상기 원문틀 데이터베이스에서 탐색한 제약조건을 입력문장에 매칭시키는 제 7 단계를 포함하는 문장골격을 기반으로 한 자동번역 방법.

청구항 6.

제 4 항 또는 제 5 항에 있어서,

상기 제 2 단계는,

상기 제약조건 매칭에 성공한 후보가 존재하는지를 확인하여, 상기 제약조건에 부가되어 있는 대역문틀 접근 포인터를 이용하여 대역문틀 데이터베이스를 탐색하는 제 8 단계;

상기 대역문틀 데이터베이스에서 탐색된 대역문틀 정보를 상기 입력문장에 적용하는 제 9 단계;

슬롯대역 데이터베이스를 탐색하여, 슬롯대역 정보를 적재하는 제 10 단계;

상기 대역문틀 적용에 성공한 모든 후보 대역문틀에 대해 각각의 문틀을 구성하는 슬롯별로 상기 슬롯대역 테이터베이스의 슬롯대역 정보를 적용하는 제 11 단계; 및

상기 슬롯대역 정보의 적용이 완료되면, 번역사전을 이용하여 원시언어 단어에 해당하는 목적언어 단어로 어휘변환을 수행하는 제 12 단계를 포함하는 문장골격을 기반으로 한 자동번역 방법.

청구항 7.

제 6 항에 있어서,

상기 제 9 단계는,

상기 대역문틀 테이터베이스의 원문 슬롯의 제약조건 스크립트를 실행하여 가중치를 계산하는 제 13 단계;

상기 원문 슬롯과 대역문 슬롯의 위치를 고려하여 문장을 구성하는 슬롯단위의 어순변환을 수행하는 제 14 단계; 및

각각의 대역문 슬롯에 대해 자질 지정 스크립트를 수행하여 해당하는 자질값을 지정하는 제 15 단계를 포함하는 문장골격을 기반으로 한 자동번역 방법.

청구항 8.

제 6 항에 있어서,

상기 제 11 단계는,

상기 슬롯대역 테이터베이스의 원문 품사에 대한 제약조건 스크립트를 수행하는 제 13 단계;

상기 제약조건 스크립트를 만족하는 원문 품사 및 대역문 품사열의 위치를 고려하여 슬롯 내부의 품사단위의 어순변환을 수행하는 제 14 단계; 및

대역문의 해당 품사에 기술된 자질 지정 스크립트를 실행하여 해당하는 자질을 지정하는 제 15 단계를 포함하는 문장골격을 기반으로 한 자동번역 방법.

청구항 9.

제 6 항에 있어서,

상기 제 3 단계는,

상기 번역사전을 이용하여 원시언어 단어에 해당하는 목적언어 단어로 어휘변환을 수행하는 제 13 단계;

상기 가중치 값을 이용하여 최종번역 결과로서 하나의 대역문틀을 선택하는 제 14 단계; 및

상기 선택된 대역문틀을 정규화된 자질값을 갖는 목적언어 형태소로 구성하는 제 15 단계를 포함하는 문장골격을 기반으로한 자동번역 방법.

청구항 10.

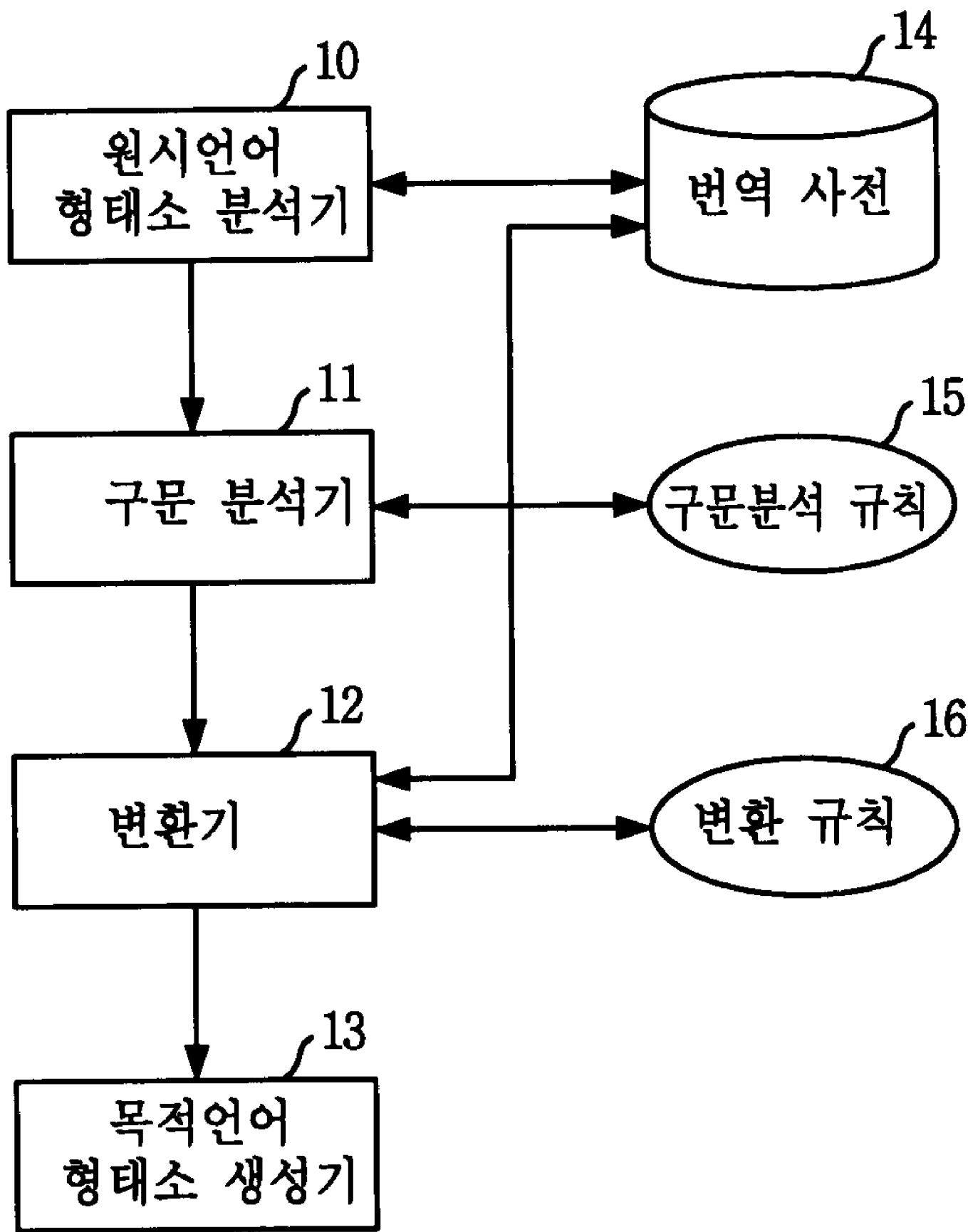
마이크로 프로세서를 구비한 자동번역 시스템에,

원시언어 형태소 분석기 및 구 단위 구문분석기로부터 구 단위 구분분석 결과를 수신하여 원문틀 데이터베이스에 저장되어 있는 제약조건을 매칭시키는 제 1 기능;

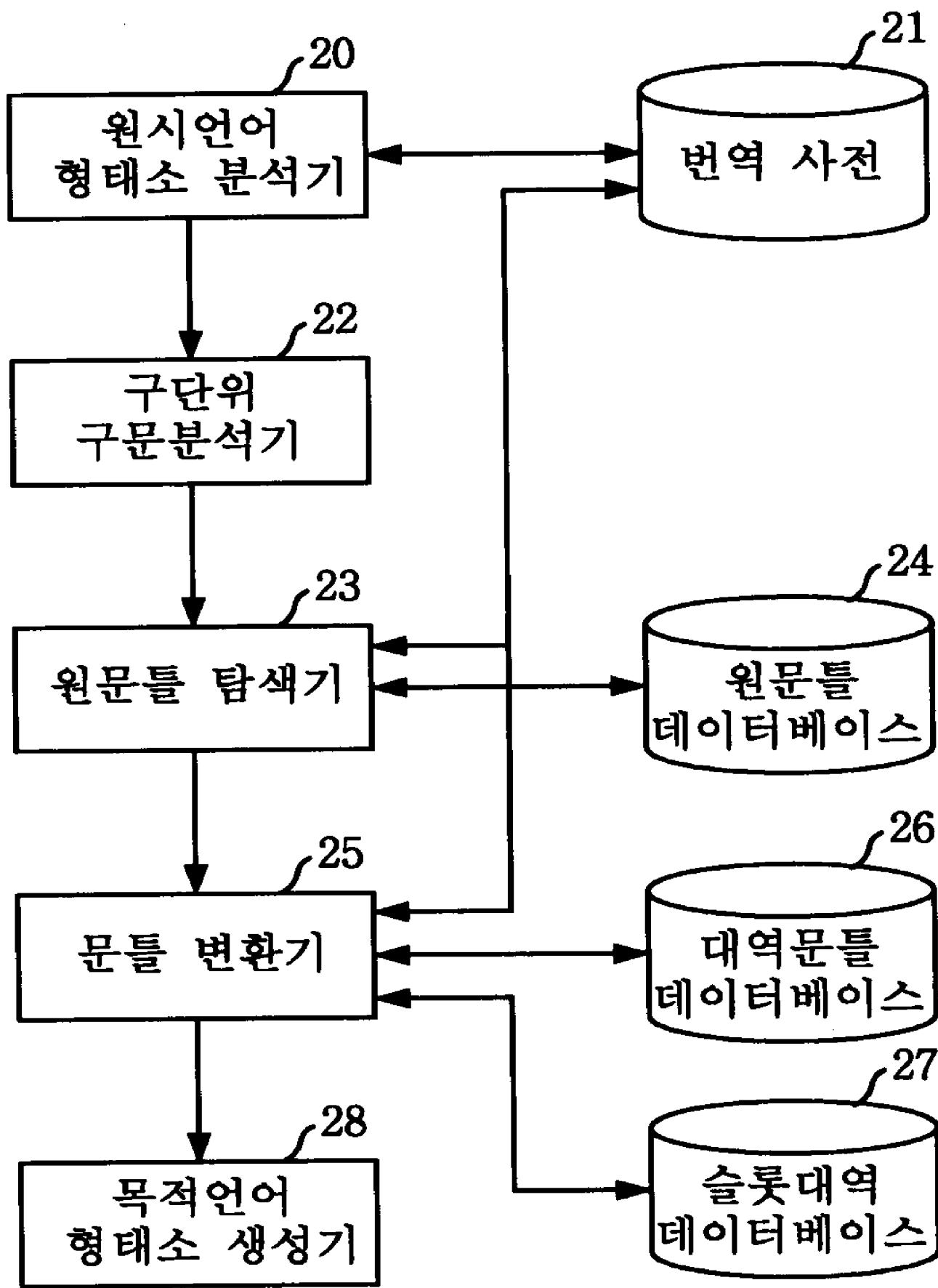
상기 제약조건이 매칭된 구문에 슬롯 단위의 어순변환 및 슬롯 내부의 품사단위의 어순변환을 수행하는 제 2 기능; 및 상기 슬롯 단위 및 품사단위로 어순 변환된 원시문장을 정규화된 자질값을 갖는 목적언어 형태소로 구성하는 제 3 기능을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

도면

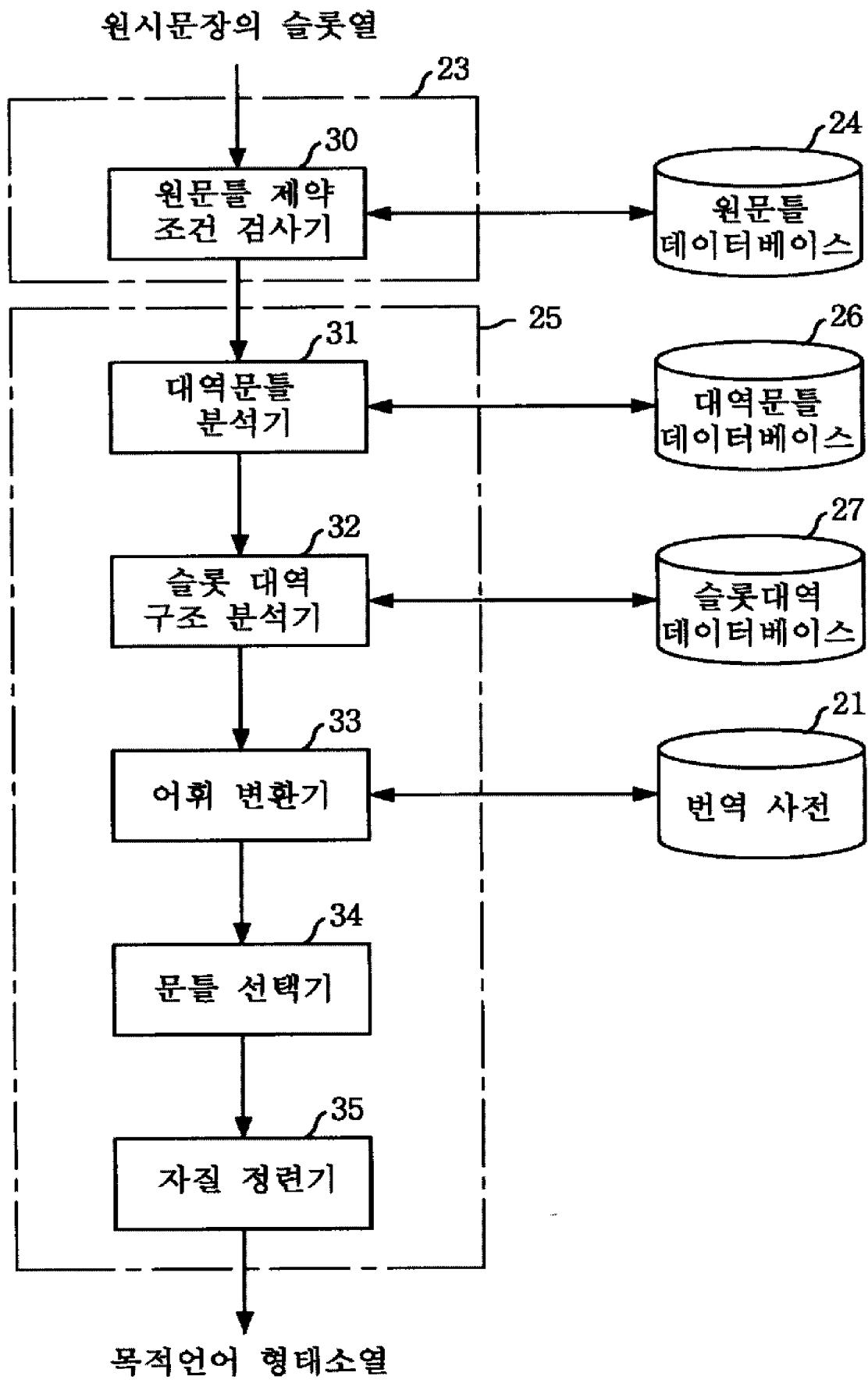
도면 1



도면 2



도면 3



도면 4

